

Combinatorial Entropy Power Inequalities: A Preliminary Study of the Stam region

Mokshay Madiman and Farhad Ghassemi

Abstract

We initiate the study of the Stam region, defined as the subset of the positive orthant in \mathbb{R}^{2^n-1} that arises from considering entropy powers of subset sums of n independent random vectors in a Euclidean space of finite dimension. We show that the class of fractionally superadditive set functions provides an outer bound to the Stam region, resolving a conjecture of A. R. Barron and the first author. On the other hand, the entropy power of a sum of independent random vectors is not supermodular in any dimension. We also develop some qualitative properties of the Stam region, showing for instance that its closure is a logarithmically convex cone.

I. INTRODUCTION

For a \mathbb{R}^d -valued random vector X with density f with respect to the Lebesgue measure on \mathbb{R}^d , the differential entropy is

$$h(X) = - \int f(x) \log f(x) dx,$$

if it exists. When X is supported on a strictly lower-dimensional set than \mathbb{R}^d (and hence does not have a density with respect to the d -dimensional Lebesgue measure), a limiting argument suggests that one should set $h(X) = -\infty$. Also, by considering for example the distribution supported on $(e, \infty) \subset \mathbb{R}$ with density

$$f(x) = \frac{1}{x(\log x)^2},$$

it is seen that the entropy can take the value $h(X) = +\infty$.

Unless explicitly mentioned, we limit ourselves to random vectors X with $h(X) < +\infty$. The entropy power of X is defined by

$$\mathcal{N}(X) = e^{2h(X)/d}.$$

Then, $\mathcal{N}(X) \in \mathbb{R}_+ := [0, \infty)$ is a non-negative real number. If a random vector X has a well-defined, finite covariance matrix K (i.e., the variance of each of the d components is finite), then it is well known that

$$h(X) \leq h(N(0, K)) = \frac{1}{2} \log[(2\pi e)^d \det(K)] < +\infty,$$

because of the fact that Gaussians maximize entropy under a constraint on the covariance matrix. Thus the class of random vectors under consideration certainly includes all those whose components have finite variances, but in fact is much larger since it also includes many heavy-tailed distributions.

There are two main motivations for considering entropy power inequalities: the first comes from the fact that it is related to probabilistic isoperimetric phenomena including the entropic central limit theorem (see, e.g., Barron [3]), and the second comes from the fact that it can be extremely useful in the study of rate and capacity regions in multi-user information theory (see, e.g., Shannon [43], Bergmans [4], Ozarow [41], Costa [11] and Oohama [40]).

Let X_1, X_2, \dots, X_n be independent random vectors. We write $[n]$ for the index set $\{1, 2, \dots, n\}$, and \emptyset for the empty set. For any nonempty $s \subset [n]$, define the subset sum

$$Y_s = \sum_{i \in s} X_i.$$

One is interested in the entropy powers $\mathcal{N}(Y_s)$ of the subset sums, which leads naturally to the following objects of study:

$$\begin{aligned} \Gamma_d(n) = \{ & [\mathcal{N}(Y_s)]_{s \subset [n]} : X_1, X_2, \dots, X_n \\ & \text{are independent } \mathbb{R}^d\text{-valued random vectors} \\ & \text{with } h(X_1 + \dots + X_n) < \infty \}. \end{aligned} \quad (1)$$

M. Madiman is with the Department of Mathematical Sciences, University of Delaware, USA. F. Ghassemi is with the Sloan School of Management, MIT, Cambridge, MA 02142, USA. Email for correspondence: madiman@udel.edu

M. Madiman is grateful for support from the U.S. National Science Foundation through the grants DMS-1409504 (CAREER) and CCF-1346564.

A preliminary version of this work [31] was presented at the 2009 IEEE International Symposium on Information Theory in Seoul, Korea. The conference paper had correctly stated results, but discussed one of them in an erroneous manner in the text; the error is corrected here.

We now give a more precise definition of the relevant objects.

Definition 1. Let $\mathbf{F}_{d,n}$ be the collection of all n -tuples $\mathbf{f} = (f_1, \dots, f_n)$ of probability density functions f_i on \mathbb{R}^d , such that if $X_i \sim f_i$ are independent, their sum has finite entropy (i.e., $h(X_1 + \dots + X_n) < \infty$). Define the set function $\nu_{\mathbf{f}} : 2^{[n]} \setminus \emptyset \rightarrow \mathbb{R}_+$ by

$$\nu_{\mathbf{f}}(s) = \mathcal{N}\left(\sum_{i \in s} X_i\right); \quad (2)$$

we will also denote $\nu_{\mathbf{f}}$ by $\nu_{(X_1, \dots, X_n)}$ when convenient. The d -dimensional Stam region is the subset of \mathbb{R}^{2^n-1} given by

$$\Gamma_d(n) = \{\nu_{\mathbf{f}} : \mathbf{f} \in \mathbf{F}_{d,n}\}.$$

The Stam region is defined by

$$\Gamma(n) = \cup_{d \in \mathbb{N}} \Gamma_d(n).$$

We name these regions after A. J. Stam in honor of his pioneering role [47] in the study of entropy power and its applications. One can extend the domain of $\nu_{\mathbf{f}}$ to the full Boolean lattice $2^{[n]}$ in a natural way by setting $\nu_{\mathbf{f}}(\emptyset) = 0$ for every $\mathbf{f} \in \mathbf{F}_{d,n}$; this would make the Stam region a subset of \mathbb{R}^{2^n} instead of \mathbb{R}^{2^n-1} , but we avoid such an extension since it is trivial.

Any inequality that relates entropy powers of different subset sums (usually called an “entropy power inequality” or EPI) gives a bound on the Stam region. Conversely, knowing the Stam region is equivalent, in principle, to knowing all EPI’s that hold and all that do not.

The objective of this note is to develop a better understanding of the Stam region. Our first result is the best outer bound known thus far for it. In particular, we show that for any dimension d , the entropy power is “fractionally superadditive”, settling a conjecture made implicitly in [30] and explicitly in [27]. We first explain what the term means.

Let \mathcal{G} be a hypergraph on $[n]$, i.e., let \mathcal{G} be a collection of nonempty subsets of $[n]$. Given \mathcal{G} , a function $\beta : \mathcal{G} \rightarrow \mathbb{R}_+$ is a *fractional partition*, if for each $i \in [n]$, we have $\sum_{s \in \mathcal{G}: i \in s} \beta_s = 1$. If there exists a fractional partition β for \mathcal{G} that is $\{0, 1\}$ -valued, then β is the indicator function for a partition of the set $[n]$ using a subset of \mathcal{G} ; hence the terminology. Note that we do not need to make reference to a hypergraph \mathcal{G} in order to define a fractional partition; the domain can always be extended to all nonempty subsets of $[n]$ by setting $\beta_s = 0$ for $s \notin \mathcal{G}$.

We say that a set function $v : 2^{[n]} \rightarrow \mathbb{R}_+$ is *fractionally superadditive* if

$$v([n]) \geq \sum_{s \in \mathcal{G}} \beta_s v(s) \quad (3)$$

holds for every fractional partition β using any hypergraph \mathcal{G} . Write $\Gamma_{FSA}(n)$ for the class of all fractionally superadditive set functions v with $v(\emptyset) = 0$.

Theorem 1. For every $n \in \mathbb{N}$, $\Gamma(n) \subset \Gamma_{FSA}(n)$.

The set function $v : 2^{[n]} \rightarrow \mathbb{R}_+$ is said to be *supermodular* if

$$v(s \cup t) + v(s \cap t) \geq v(s) + v(t)$$

for all sets $s, t \subset [n]$. Write $\Gamma_{SM}(n)$ for the class of all supermodular set functions v with $v(\emptyset) = 0$.

It is known [39], [37] that $\Gamma_{SM}(n) \subsetneq \Gamma_{FSA}(n)$. In [27], it was asked whether in fact the entropy power is supermodular, i.e., whether the set function $\nu_{\mathbf{f}} \in \Gamma_{SM}(n)$ for every $\mathbf{f} \in \mathbf{F}_{d,n}$. Our second result is to show that the answer to this question is no in general.

Theorem 2. For any $n \geq 3$, $\Gamma(n) \not\subset \Gamma_{SM}(n)$.

The Stam region has some pleasing geometric properties. Unfortunately we have not yet been able to determine if the closure of the Stam region is convex; however it does have some restricted convexity properties. We say that a set $A \subset \mathbb{R}_+^k$ is *logarithmically convex* if for any two points $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$ in A , the point $(x_1^\lambda y_1^{1-\lambda}, \dots, x_k^\lambda y_k^{1-\lambda}) \in A$ for each $\lambda \in [0, 1]$. We say that a set $A \subset \mathbb{R}^k$ is *orthogonally convex* if any line segment parallel to any of the coordinate axes connecting two points of A lies totally within A . It is trivial to check that any subset of \mathbb{R}_+^k that is logarithmically convex is necessarily orthogonally convex.

Theorem 3. The closure $\overline{\Gamma(n)}$ of the Stam region $\Gamma(n)$ is a logarithmically convex cone in \mathbb{R}^{2^n-1} . Consequently, $\overline{\Gamma(n)}$ is orthogonally convex.

For the case where one has only two random variables, we can give a complete description of the Stam region.

Theorem 4. $\overline{\Gamma(2)} = \overline{\Gamma_1(2)} = \Gamma_{FSA}(2)$. In particular, $\overline{\Gamma(2)}$ is a closed, convex, polyhedral cone in \mathbb{R}_+^3 .

The general problem is significantly more complicated.

Theorem 5. For $n \geq 3$, $\overline{\Gamma(3)} \subsetneq \Gamma_{FSA}(3)$.

The closure of the Stam region, which features in the preceding three theorems, is in fact closely related to the “Stam region” involving entropy rates of a class of stochastic processes. By a \mathbb{R} -valued stochastic process, we will mean a discrete-time stochastic process $X = (X_1, X_2, \dots)$, with each X_i taking values in \mathbb{R} . (More precisely, we may interpret this process as a measurable function from the underlying probability space to the set $\mathbb{R}^{\mathbb{N}}$ equipped with the cylindrical σ -algebra.) Recall that the *entropy rate* of a \mathbb{R} -valued stochastic process is defined by

$$\bar{h}(X) = \lim_{d \rightarrow \infty} \frac{h(X_1, \dots, X_d)}{d},$$

if the limit exists.

Definition 2. Let \mathbf{F}_n be the collection of all n -tuples $\mathbf{X} = (X^{(1)}, \dots, X^{(n)})$ of independent \mathbb{R} -valued stochastic processes, such that each $X^{(i)}$ as well as their sum has finite entropy rate (i.e., $\bar{h}(X^{(1)} + \dots + X^{(n)}) < \infty$). Define the set function $\nu_{\mathbf{X}} : 2^{[n]} \setminus \emptyset \rightarrow \mathbb{R}_+$ by

$$\nu_{\mathbf{X}}(s) = \exp \left\{ 2\bar{h} \left(\sum_{i \in s} X^{(i)} \right) \right\}. \quad (4)$$

The ∞ -dimensional Stam region is the subset of $\mathbb{R}^{2^n - 1}$ given by

$$\Gamma_{\infty}(n) = \{\nu_{\mathbf{X}} : \mathbf{X} \in \mathbf{F}_n\}.$$

Now we can relate the finite-dimensional and infinite-dimensional Stam regions.

Theorem 6. For any $n \in \mathbb{N}$, $\overline{\Gamma(n)} = \Gamma_{\infty}(n)$.

This gives a pleasing “physical” interpretation of the closure of the Stam region— the closure can be thought of not just as a topological operation on a set that has intrinsic probabilistic motivation, but as an extension of the definition of the Stam region to random vectors of dimension ∞ .

Let us note that the restriction $h(X_1 + \dots + X_n) < \infty$ in the definition of the Stam region is not essential. One can define the *extended d -dimensional Stam region* as

$$\tilde{\Gamma}_d(n) = \{\nu_{\mathbf{f}} : \mathbf{f} \in \tilde{\mathbf{F}}_{d,n}\},$$

where $\tilde{\mathbf{F}}_{d,n}$ is the collection of all n -tuples $\mathbf{f} = (f_1, \dots, f_n)$ of probability density functions f_i on \mathbb{R}^d such that $h(f_i)$ exists for each $i \in [n]$. The *extended Stam region* would then be defined by

$$\tilde{\Gamma}(n) = \cup_{d \in \mathbb{N}} \tilde{\Gamma}_d(n);$$

both $\tilde{\Gamma}_d(n)$ and $\tilde{\Gamma}(n)$ are then subsets of $\mathbb{R}_+^{2^n}$, with the extended nonnegative real numbers $\mathbb{R}_+ = [0, \infty]$ replacing \mathbb{R}_+ . All our theorems, with straightforward minor modifications, can be stated for the extended Stam regions.

After developing some combinatorial preliminaries in Section II, we prove the above theorems in Section III. Although Theorem 2 asserts that supermodularity is false for the Stam region, the proof given in Section III does not shed light on the whether supermodularity might in fact hold in the one-dimensional case (real-valued random variables); this question is discussed in Section IV. Section V contains some additional remarks of possible interest.

II. COMBINATORIAL PRELIMINARIES

The preliminaries we will need range from a lemma about fractional superadditivity that we could not find explicitly stated in the literature, to some facts about supermodular functions on lattices.

We now state and prove a useful lemma about fractional superadditivity that is at the heart of our proof of Theorem 1. We will use in the proof the following standard fact about a vertex or extreme point of a polytope: a point in a polytope is an extreme point if and only if it is the unique meeting point of several faces (i.e., a set of faces intersects in a singleton, containing that point). Indeed, if it were not unique, then it could not be an extreme point, since there would be a line segment in the polytope containing it.

Proposition 1. [A SUFFICIENT CONDITION FOR FRACTIONAL SUPERADDITIVITY] Consider a set function $v : 2^{[n]} \rightarrow \mathbb{R}_+$. Let \mathcal{G} be a r -regular multihypergraph on $[n]$, i.e., let \mathcal{G} be a collection of subsets of $[n]$ (possibly repeated), such that every index i lies in exactly r of the elements of \mathcal{G} . Suppose v satisfies, for any $r \in \mathbb{N}$,

$$v([n]) \geq \frac{1}{r} \sum_{s \in \mathcal{G}} v(s)$$

where \mathcal{G} is any r -regular multihypergraph on $[n]$. Then,

$$v([n]) \geq \sum_{s \in \mathcal{G}} \beta_s v(s) \quad (5)$$

holds for every fractional partition β using any multihypergraph \mathcal{G} on $[n]$.

Proof. Denote by $\mathbf{1}_s$ the indicator function of the subset s defined on the domain $[n]$, i.e., $\mathbf{1}_s(i) = 1$ if $i \in s$, and $\mathbf{1}_s(i) = 0$ otherwise. Consider the space of all fractional partitions on $[n]$, i.e.,

$$\mathcal{B} = \left\{ \beta : 2^{[n]} \setminus \phi \rightarrow \mathbb{R}_+ \left| \sum_{s \subset [n] \setminus \phi} \beta_s \mathbf{1}_s = \mathbf{1}_{[n]} \right. \right\}.$$

Clearly, \mathcal{B} can be viewed as a subset of the Euclidean space of dimension $2^n - 1$ (since each point of it is defined by $2^n - 1$ real numbers). Furthermore, $\mathcal{B} = \mathcal{B}' \cap \mathcal{O}_+$, where

$$\mathcal{B}' = \left\{ \beta : 2^{[n]} \setminus \phi \rightarrow \mathbb{R} \left| \sum_{s \subset [n] \setminus \phi} \beta_s \mathbf{1}_s = \mathbf{1}_{[n]} \right. \right\}$$

is an affine subspace of dimension $2^n - 1 - n$ and $\mathcal{O}_+ = \{\beta | \beta_s \geq 0 \ \forall s \in 2^{[n]} \setminus \phi\}$ is the closed positive orthant. In particular, \mathcal{B} is a non-empty, compact, convex set (in fact, a closed polytope), so that by the Krein-Milman theorem, \mathcal{B} is the convex hull of its extreme points. Thus to prove (5) for every $\beta \in \mathcal{B}$, it is sufficient to prove (5) for every $\beta \in \text{Ex}(\mathcal{B})$, where $\text{Ex}(\mathcal{B})$ denotes the set of extreme points of \mathcal{B} .

For a fractional partition β , its *support* is defined as the collection of subsets s of $[n]$ such that $\beta_s > 0$. Given a hypergraph \mathcal{G} , the set of fractional partitions supported by \mathcal{G} is clearly the set of positive solutions of the linear equation

$$M_{\mathcal{G}} \beta = \mathbf{1}, \quad (6)$$

where $M_{\mathcal{G}}$ is the $n \times |\mathcal{G}|$ 0-1 matrix defined by $M_{i,s} = \mathbf{1}_{i \in s}$ for $i \in [n], s \in \mathcal{G}$, and $\mathbf{1}$ is the column vector in \mathbb{R}^n consisting of all ones. In general, the number of fractional partitions supported by \mathcal{G} could be either 0, 1 or infinite, depending on \mathcal{G} .

Every face of the polytope \mathcal{B} corresponds to one of the inequality constraints being tight (i.e., $\beta_s = 0$ for at least one set s). Given an extreme point β of the polytope \mathcal{B} , we know it is the *unique* meeting point of several faces; let these faces correspond to setting $\beta_s = 0$ for s lying in the collection of sets \mathcal{G}' . Then the complement \mathcal{G} of \mathcal{G}' is the support of β , and β must be the unique fractional partition supported by \mathcal{G} . By the previous paragraph, we know this means that β is the unique, strictly positive solution of the equation (6). Consequently one must have $|\mathcal{G}| \leq n$, with precisely $|\mathcal{G}|$ of the n rows of $M_{\mathcal{G}}$ being linearly independent, so that $M_{\mathcal{G}}$ has full rank. Since $M_{\mathcal{G}}$ has only integer entries, a little bit of thought shows that performing Gaussian elimination to solve (6) will result in its unique solution having only rational entries. Thus we have deduced that any fractional partition in $\text{Ex}(\mathcal{B})$ has rational¹ coefficients.

By writing all the coefficients of β with a common denominator, one sees that (5) may be written as

$$v_n([n]) \geq \frac{1}{R} \sum_{s \in \mathcal{G}} c_s v_n(s),$$

where c_s is a positive integer. One may write this as

$$v_n([n]) \geq \frac{1}{R} \sum_{s \in \mathcal{G}''} v_n(s),$$

where \mathcal{G}'' is the multihypergraph with c_s copies of the set s . Note that \mathcal{G}'' is clearly R -regular. □

Remark 1. The argument we use to characterize $\text{Ex}(\mathcal{B})$ is similar to that in Gill and Grünwald [19] and seems also implicit in Friedgut and Kahn [16]; it likely goes back to the early literature on cooperative game theory (see, e.g., Shapley [44]), although we were unable to find the precise statement we wanted in these early references. Incidentally a fractional partition is interpreted in [18], [19] as a “coarsening at random” or CAR mechanism; this is a probabilistic rule that replaces any point x in the ground set $[n]$ with a subset A of $[n]$ containing x , in such a way that the probability of observing A is the same for all x that are contained in A .

The last observation we need concerns supermodular functions; first we need a definition of supermodularity for functions on \mathbb{R}^n and not just set functions.

¹Another way of seeing this would be to observe that $\beta = M_{\mathcal{G}}^+ \mathbf{1}$, where $M_{\mathcal{G}}^+$ is the Moore-Penrose pseudo-inverse of M . Then the rationality of entries of $M_{\mathcal{G}}$ implies that of the entries of $M_{\mathcal{G}}^+$, and hence the rationality of β .

Definition 3. A function $f : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is supermodular if

$$f(x) + f(y) \leq f(x \vee y) + f(x \wedge y)$$

for any $x, y \in \mathbb{R}_+^n$, where $x \vee y$ denotes the componentwise maximum of x and y and $x \wedge y$ denotes the componentwise minimum of x and y .

The fact that supermodular functions are closely related to functions with increasing differences is classical (see, e.g., [49], which describes more general results involving arbitrary lattices), but we give the proof for completeness.

Proposition 2. Suppose a function $f : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is in C^2 , i.e., it is twice-differentiable with a continuous Hessian matrix. Then f is supermodular if and only if

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} \geq 0$$

for every distinct $i, j \in [n]$, and for any $x \in \mathbb{R}_+^n$.

Proof. Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is supermodular. Let $i, j \in [n]$ with $i < j$, and pick a set of coordinates a_k for each $k \in A$, where $A = [n] \setminus \{i, j\}$. For any $x \in \mathbb{R}^n$ with $x_k = a_k$ for each $k \in A$, set

$$g(x_i, x_j) = f(x).$$

Then for $y_i < z_i$ and $y_j < z_j$,

$$\begin{aligned} g(y_i, z_j) - g(y_i, y_j) &= g(y_i, z_j) - g(y_i \wedge z_i, y_j \wedge z_j) \\ &\leq g(y_i \vee z_i, y_j \vee z_j) - g(z_i, y_j) \\ &= g(z_i, z_j) - g(z_i, y_j); \end{aligned}$$

in other words, differences of g (and hence of f) in the j -th coordinate are increasing in the i -th coordinate. If f is C^2 , this implies that

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} \geq 0$$

on the interior of the domain. Since the same argument can be repeated for each pair i, j of distinct indices, we obtain non-negativity of all off-diagonal entries of the Hessian matrix of f .

Now suppose

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} \geq 0$$

for any $x \neq 0$. By standard calculus, this implies that f has increasing differences, in the sense that for any pair $i \neq j$ of distinct indices, and for every $a = (a_k : k \in A)$ where $A = [n] \setminus \{i, j\}$, one has that differences in the j -th coordinate, namely

$$f(a, x_i, x_j) - f(a, x_i, \tilde{x}_j)$$

with $x_j > \tilde{x}_j$, are increasing as functions of x_i . (We have abused notation for convenience here since f may not be symmetric in its arguments, but we are implicitly assuming that x_i is the i -th argument and x_j is the j -th argument when we write $f(a, x_i, x_j)$, etc.) Thus we have

$$\begin{aligned} f(y) - f(y \wedge z) &= \sum_{i \in [n]} \left[f(y_1^{i-1}, y_i, y_{i+1}^n \wedge z_{i+1}^n) - f(y_1^{i-1}, y_i \wedge z_i, y_{i+1}^n \wedge z_{i+1}^n) \right] \\ &\leq \sum_{i \in [n]} \left[f(y_1^{i-1} \vee z_1^{i-1}, y_i, z_{i+1}^n) - f(y_1^{i-1} \vee z_1^{i-1}, y_i \wedge z_i, z_{i+1}^n) \right], \end{aligned}$$

where the inequality follows from the increasing differences property of f . Now we can apply the trivial fact that for any function g on the real line, one has $g(x) + g(y) = g(x \vee y) + g(x \wedge y)$ to deduce that the last expression equals

$$\sum_{i \in [n]} \left[f(y_1^{i-1} \vee z_1^{i-1}, y_i \vee z_i, z_{i+1}^n) - f(y_1^{i-1} \vee z_1^{i-1}, z_i, z_{i+1}^n) \right],$$

which telescopes down to $f(y \vee z) - f(z)$. Combining the preceding two displays, we have proved that

$$f(y) - f(y \wedge z) \leq f(y \vee z) - f(z),$$

which is precisely supermodularity of f . \square

Our final very simple lemma connects supermodularity for functions defined on \mathbb{R}_+^n to supermodularity for set functions. For a set $s \subset [n]$, we use $e(s)$ to denote the vector in \mathbb{R}^n such that for each $i \in [n]$, the i -th coordinate of $e(s)$ is 1 if and only if $i \in s$ (or in other words, $e_i(s) := \mathbf{1}_s(i)$).

Lemma 1. *If $f : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is supermodular, and we set $\bar{f}(s) := f(e(s))$ for each $s \subset [n]$, then \bar{f} is a supermodular set function.*

Proof. Observe that

$$\begin{aligned} \bar{f}(s \cup t) + \bar{f}(s \cap t) &= f(e(s \cup t)) + f(e(s \cap t)) \\ &= f(e(s) \vee e(t)) + f(e(s) \wedge e(t)) \\ &\geq f(e(s)) + f(e(t)) \\ &= \bar{f}(s) + \bar{f}(t). \end{aligned}$$

□

III. PROOFS

A. Proof of Theorem 1

In [30], the following EPI for an arbitrary multihypergraph \mathcal{G} on $[n]$ was demonstrated. If r is the maximum number of subsets in \mathcal{G} in which any one index i can appear, for $i = 1, \dots, n$, then

$$\mathcal{N}(X_1 + \dots + X_n) \geq \frac{1}{r} \sum_{s \in \mathcal{G}} \mathcal{N}\left(\sum_{j \in s} X_j\right). \quad (7)$$

In other words, for any $\mathbf{f} \in \mathbf{F}_{d,n}$, $\nu_{\mathbf{f}}$ satisfies the assumption in Proposition 1, and hence we have that $\nu_{\mathbf{f}}$ is fractionally superadditive.

B. Proof of Theorem 2

Observe that if the function $\nu_{\mathbf{f}}(s) = \mathcal{N}(\sum_{i \in s} X_i)$ were supermodular, then specializing to multivariate Gaussians would imply a similar supermodularity for the d -th root of the determinant of sums of positive definite matrices. To be precise, consider the set function

$$\nu_G(s) = \det\left(\sum_{k \in s} M_k\right)^{\frac{1}{d}}, \quad s \subset [n],$$

where M_1, M_2, \dots, M_n are $d \times d$ positive definite matrices. (Here we use ν_G to indicate that this is the function $\nu_{\mathbf{f}}$ specialized to Gaussians, i.e., when f_i is the density of the $N(0, M_i)$ distribution.) We will show that ν_G is not supermodular for $d = 2$.

First proof. We exhibit a numerical counterexample. Consider

$$A = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}.$$

Then it is easily seen that

$$\begin{aligned} \det(A)^{1/2} &= 1, \quad \det(A + B + C)^{1/2} = 2.5 + \epsilon \\ \det(A + B)^{1/2} &= 2.5, \quad \det(A + C)^{1/2} = \sqrt{1 + 2.5\epsilon + \epsilon^2} \end{aligned}$$

Observe that when ϵ is small, $\det(A + C)^{1/2} = 1 + 1.25\epsilon + o(\epsilon)$, so that we can arrange for $\det(A + B)^{1/2} + \det(A + C)^{1/2} = 3.5 + 1.25\epsilon + o(\epsilon)$ to exceed $\det(A)^{1/2} + \det(A + B + C)^{1/2} = 3.5 + \epsilon$.

Second proof. We present a second, more complicated, proof in an attempt to give some additional insight into why supermodularity fails. To prove that ν_G is not supermodular, we first show that its continuous analogue is not supermodular (in the continuous sense). In other words, let $v : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ be defined by

$$v(x) = \det^{\frac{1}{d}}\left(\sum_{k \in [n]} x_k M_k\right),$$

where M_1, M_2, \dots, M_n are $d \times d$ positive definite matrices. We will show that the function v is not supermodular, i.e., there are $x, x' \in \mathbb{R}_+^n$ such that the following inequality is violated

$$v(x) + v(x') \leq v(x \vee x') + v(x \wedge x') \quad (8)$$

where $x \vee x'$ denotes the componentwise maximum and $x \wedge x'$ denotes the componentwise minimum of x and x' .

To show that (8) is violated, it suffices to show, thanks to Proposition 2, that v is C^2 (which is immediate from the fact that the determinant of a linear combination of matrices is a polynomial in the coefficients) and that

$$\frac{\partial^2 v(x)}{\partial x_j \partial x_i} < 0 \quad (9)$$

for some $x \in \mathbb{R}_+^n$. Setting $M = \sum_{k \in [n]} x_k M_k$, we note that

$$\frac{\partial^2 v(x)}{\partial x_j \partial x_i} = \frac{1}{d} (\det M)^{\frac{1}{d}} \times \left[\frac{1}{d} \text{tr} \{M^{-1} M_i\} \text{tr} \{M^{-1} M_j\} - \text{tr} (M^{-1} M_j M^{-1} M_i) \right]. \quad (10)$$

However, it is easy to show (see, e.g., Zhang [55], page 166) that there are $d \times d$ positive definite matrices A and B for which

$$\frac{1}{d} \text{tr}(A) \text{tr}(B) < \text{tr}(AB).$$

Hence, the last term in (10) can be negative. As a numerical example, consider $d = 2$ and

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ 3 & 7 \end{bmatrix}.$$

It then holds that $\frac{1}{2} \text{tr}(A) \text{tr}(B) = 18$ whereas $\text{tr}(AB) = 19$.

Finally, note that the violation of supermodularity in the discrete domain follows from this result of violation of supermodularity in continuous domain. Indeed, we have shown that there is $x \in \mathbb{R}_+^n$ such that (9) is true. It then follows that there are $s_i, s_j > 0$ such that

$$\begin{aligned} & v(x_1, \dots, x_i + s_i, \dots, x_j, \dots, x_n) + v(x_1, \dots, x_i, \dots, x_j + s_j, \dots, x_n) \\ & > v(x_1, \dots, x_i + s_i, \dots, x_j + s_j, \dots, x_n) + v(x). \end{aligned}$$

Consider $A = s_i M_i$, $B = s_j M_j$, $C = \sum_{k \in [n]} x_k M_k$. Evidently, A, B, C are positive definite. However, according to the above inequality and the definition of the function v , we have shown that

$$\begin{aligned} & \det(A + C)^{\frac{1}{d}} + \det(B + C)^{\frac{1}{d}} \\ & > \det(A + B + C)^{\frac{1}{d}} + \det(C)^{\frac{1}{d}}, \end{aligned}$$

so that neither ν_G nor ν is supermodular.

C. Proof of Theorem 3

The fact that each $\Gamma_d(n)$ is a cone follows just from scaling. By definition, any point $z \in \Gamma_d(n)$ is just $\nu_{(X_1, \dots, X_n)}$ for some independent random vectors X_1, \dots, X_n (each of which is in \mathbb{R}^d). For any $\lambda > 0$, the point λz is then simply $\nu_{(\sqrt{\lambda} X_1, \dots, \sqrt{\lambda} X_n)}$ and hence also lies in $\Gamma_d(n)$. Since each $\Gamma_d(n)$ is a cone, so are $\Gamma(n)$ and $\overline{\Gamma(n)}$.

What remains is to show that $\overline{\Gamma(n)}$ is a logarithmically convex set. To do this, we will need the following lemma.

Lemma 2. *For any given $d, d' \in \mathbb{N}$, the set*

$$Q_{d,d'} = \left\{ \frac{md}{md + ld'} : m, l \in \mathbb{Z}_+ \right\}$$

is a dense subset of $[0, 1]$.

Proof. Letting $z(m, l) = \frac{md}{md + ld'}$, the image $Q_{d,d'}$ of $\mathbb{Z}_+ \times \mathbb{Z}_+$ under z is clearly contained in the interval $[0, 1]$. Since $[0, 1]$ is closed, we also have $\overline{Q_{d,d'}} \subset [0, 1]$; thus what remains to be proved is that every point of $[0, 1]$ is in the closure of $Q_{d,d'}$.

Let x be an arbitrary rational number in $[0, 1]$; then x is of the form $\frac{p}{p+q}$, where $p, q \in \mathbb{Z}_+$. Writing

$$z(m, l) = \frac{d}{d + \frac{l}{m} d'} = \frac{1}{1 + \frac{l}{m} \cdot \frac{d'}{d}},$$

we see that x is in the image of $Q_{d,d'}$ of $\mathbb{Z}_+ \times \mathbb{Z}_+$ under z by taking $l = qd$ and $m = pd'$. In other words, we have $\mathbb{Q} \cap [0, 1] \subset Q_{d,d'}$, which—by the density of the rationals in the reals—implies that $[0, 1] \subset \overline{Q_{d,d'}}$.

Thus it is demonstrated that $\overline{Q_{d,d'}} = [0, 1]$. □

Remark 2. Lemma 2 is related to interesting problems in number theory, such as the coin problem (sometimes called the diophantine Frobenius problem), which goes back to 1882 and asks for the largest integer that cannot be written as a nonnegative linear combination of a given set of natural numbers. (A theorem of Schur ensures that there is such a largest integer.) The book of Ramírez Alfonsín [42] contains many more details.

To show that $\overline{\Gamma(n)}$ is a logarithmically convex set, let us start with two points in $\Gamma(n)$, say $\nu_{\mathbf{X}} := \nu_{(X_1, \dots, X_n)} \in \Gamma_d(n)$ and $\nu_{\mathbf{Y}} := \nu_{(Y_1, \dots, Y_n)} \in \Gamma_{d'}(n)$, where our notation is as explained in Definition 1. Let $\mathbf{Z}^{(m,l)}$ be the vector formed by stacking m independent copies of \mathbf{X} with l independent copies of \mathbf{Y} ; in other words,

$$\mathbf{Z}^{(m,l)} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(m)} \\ \mathbf{Y}^{(1)} \\ \vdots \\ \mathbf{Y}^{(l)} \end{pmatrix} = \begin{pmatrix} X_1^{(1)} & \dots & X_n^{(1)} \\ \vdots & \dots & \vdots \\ X_1^{(m)} & \dots & X_n^{(m)} \\ Y_1^{(1)} & \dots & Y_n^{(1)} \\ \vdots & \dots & \vdots \\ Y_1^{(l)} & \dots & Y_n^{(l)} \end{pmatrix},$$

where each $X_i^{(k)} \in \mathbf{R}^d$ and each $Y_i^{(k)} \in \mathbf{R}^{d'}$. Thus, denoting the i -th column of $\mathbf{Z}^{(m,l)}$ by $Z_i^{(m,l)}$, we see that each $Z_i^{(m,l)}$ is a random vector in $\mathbf{R}^{md+ld'}$. For each $s \subset [n]$, we have

$$\begin{aligned} \nu_{(Z_1^{(m,l)}, \dots, Z_n^{(m,l)})}(s) &= \mathcal{N}\left(\sum_{i \in s} Z_i^{(m,l)}\right) \\ &= \mathcal{N}\left(\sum_{i \in s} X_i^{(1)}, \dots, \sum_{i \in s} X_i^{(m)}, \sum_{i \in s} Y_i^{(1)}, \dots, \sum_{i \in s} Y_i^{(l)}\right) \\ &= \mathcal{N}\left(\sum_{i \in s} X_i\right)^{\frac{md}{md+ld'}} \mathcal{N}\left(\sum_{i \in s} Y_i\right)^{\frac{ld'}{md+ld'}} \\ &= \nu_{\mathbf{X}}(s)^\lambda \nu_{\mathbf{Y}}(s)^{1-\lambda}, \end{aligned}$$

where the third equality follows from the assumed independence and the fact that the dimension appears in the exponent in the definition of entropy power.

Thus we have shown that if $a \in \Gamma_d(n)$ and $b \in \Gamma_{d'}(n)$, then

$$\{a^\lambda b^{1-\lambda} : \lambda \in Q_{d,d'}\}$$

is a subset of $\Gamma(n)$. Invoking the density of $Q_{d,d'}$ in $[0, 1]$ (Lemma 2) and the continuity of the map $\lambda \mapsto a^\lambda b^{1-\lambda}$, we see that

$$\{a^\lambda b^{1-\lambda} : \lambda \in [0, 1]\}$$

is a subset of $\overline{\Gamma(n)}$, which is precisely the claimed logarithmic convexity.

D. Proof of Theorem 4

Since $\Gamma_{FSA}(2) = \{(u_{\{1\}}, u_{\{2\}}, u_{\{1,2\}}) \in \mathbb{R}_+^3 : u_{\{1,2\}} \geq u_{\{1\}} + u_{\{2\}}\}$, it suffices to show that for any given $u_{\{1\}}, u_{\{2\}} \in \mathbb{R}_+$, and for a dense subset of $u_{\{1,2\}}$ such that $u_{\{1,2\}} \geq u_{\{1\}} + u_{\{2\}}$, there exist independent random vectors X, Y such that $\mathcal{N}(X) = u_{\{1\}}$, $\mathcal{N}(Y) = u_{\{2\}}$, and $\mathcal{N}(X + Y) = u_{\{1,2\}}$.

We find it useful to adapt a construction of Bobkov and Chistyakov [7, Example 3] for our purposes. Consider the uniform density f on a set

$$A = \cup_{n \in \mathbb{N}} (2^n, 2^n + a_n),$$

where $a_n \geq 0$ for each n , and $\sum_{n \in \mathbb{N}} a_n = 1$. Alternatively, one can write f in the form

$$f(x) = \sum_{n \in \mathbb{N}} a_n f_n(x),$$

where $f_n(x) = a_n^{-1} \mathbf{1}_{(2^n, 2^n + a_n)}(x)$. If X, X' are i.i.d. and have density f , then $h(X) = 0$ since A has length 1, and [7] showed that

$$\begin{aligned} h(X + X') &= -2 \log 2 + 2 \log 2 \sum_{n \in \mathbb{N}} s_n a_n + \sum_{n \in \mathbb{N}} \left(n - \frac{1}{2}\right) a_n^2 \\ &\quad + \sum_{n \in \mathbb{N}} a_n^2 \log \frac{1}{a_n} + 2 \sum_{n=2}^{\infty} s_{n-1} a_n \log \frac{1}{a_n}, \end{aligned}$$

where $s_n = a_1 + \dots + a_n$. By choosing the weights a_n appropriately, one can clearly make $h(X + X')$ larger than any constant one might pick.² For example, for N chosen large enough, one may take $a_n = 1/N$ for $1 \leq n \leq N$, and $a_n = 0$ for $n > N$. By scaling (i.e., considering the random variables $Y = bX$ and $Y' = bX'$ for $b > 0$, one has examples where $h(Y) = \log b$ is any specified real number, and $h(Y + Y')$ is larger than an arbitrary specified constant.

Now suppose $u_{\{1\}}, u_{\{2\}}$ are given, and assume without loss of generality that $u_{\{1\}} \leq u_{\{2\}}$. Fix an arbitrarily large positive constant C . By the preceding paragraph, there exists a random vector X such that $h(X) = s := \frac{d}{2} \log u_{\{1\}}$, and $\mathcal{N}(X + X') > C$, where X' is an independent copy of X . Observe that the function $u \mapsto h(X' + \sqrt{u}Z)$, where Z is an independent standard Gaussian, maps $[0, \infty)$ to $[s, \infty)$, and is also smooth and monotonically increasing (as quantified by the classical de Bruijn identity). Consequently, there must exist a $u > 0$ such that $h(X' + \sqrt{u}Z) = t := \frac{d}{2} \log u_{\{2\}}$. Setting $Y = X' + \sqrt{u}Z$, observe that $\mathcal{N}(X) = u_{\{1\}}$, $\mathcal{N}(Y) = u_{\{2\}}$, and $\mathcal{N}(X + Y) = \mathcal{N}(X + X' + \sqrt{u}Z) \geq \mathcal{N}(X + X') > C$.

From the preceding paragraph, for any given $u_{\{1\}}, u_{\{2\}} > 0$, and for arbitrary C , there exists $C' > C$ such that $(u_{\{1\}}, u_{\{2\}}, C') \in \Gamma_1(2) \subset \overline{\Gamma(2)}$. On the other hand, by considering one-dimensional Gaussians of appropriate variances, clearly $(u_{\{1\}}, u_{\{2\}}, u_{\{1\}} + u_{\{2\}}) \in \Gamma_1(2) \subset \overline{\Gamma(2)}$. By Theorem 3, $\overline{\Gamma(2)}$ is a convex cone, from which it follows that the ray

$$R_{u_{\{1\}}, u_{\{2\}}} = \{(u_{\{1\}}, u_{\{2\}}, u_{\{1,2\}}) : u_{\{1,2\}} \geq u_{\{1\}} + u_{\{2\}}\}$$

is a subset of $\overline{\Gamma(2)}$.

We wish to now extend this subset relation to the one remaining circumstance—namely when either $u_{\{1\}}$ or $u_{\{2\}}$ (or both) is 0. This can be treated by extending another example of Bobkov and Chistyakov [7, Example 1]. For every $\epsilon > 0$, consider the density

$$f_\epsilon(x) = \frac{1}{x(\log \frac{1}{x})^{1+\epsilon}}, \quad x \in (0, \frac{1}{e}).$$

The entropy of f_ϵ exists for every $\epsilon > 0$, and if X is drawn from f_ϵ , then $h(X) > -\infty$ if and only if $\epsilon > 1$. Furthermore, if Y is drawn from $f_{\epsilon'}$ for some $\epsilon' > 0$, then the density of $X + Y$ behaves near 0 like $f_{\epsilon+\epsilon'}$, so that $h(X + Y) > -\infty$ if and only if $\epsilon + \epsilon' > 1$. Now simply by scaling, i.e., considering aX for $a > 0$, one can generate examples where $u_{\{1\}} = u_{\{2\}} = 0$, and $u_{\{1,2\}}$ is any positive real number.

We have now proved that $R_{u_{\{1\}}, u_{\{2\}}} \subset \overline{\Gamma(2)}$ for arbitrary $u_{\{1\}}, u_{\{2\}} \geq 0$. Since

$$\Gamma_{FSA}(2) = \cup_{u_{\{1\}}, u_{\{2\}} \geq 0} R_{u_{\{1\}}, u_{\{2\}}},$$

we deduce that $\Gamma_{FSA}(2) \subset \overline{\Gamma(2)}$. Since the opposite inclusion follows from Theorem 1 and the fact that $\Gamma_{FSA}(2)$ is closed, we have completed the proof of the fact that $\Gamma_{FSA}(2) = \overline{\Gamma(2)}$.

E. Proof of Theorem 5

Fix $a, b, c > 0$. Consider the sets

$$\begin{aligned} R_{a,b,c} = \{ & u_{\{1,2,3\}} : u = (u_{\{1\}}, u_{\{2\}}, u_{\{3\}}, u_{\{1,2\}}, u_{\{2,3\}}, u_{\{1,3\}}, u_{\{1,2,3\}}) \in \Gamma(3), \\ & u_{\{1\}} = a, u_{\{2\}} = b, u_{\{3\}} = c, \\ & u_{\{1,2\}} = a + b, u_{\{2,3\}} = b + c, u_{\{1,3\}} = c + a \} \end{aligned}$$

and

$$\begin{aligned} R'_{a,b,c} = \{ & u_{\{1,2,3\}} : u = (u_{\{1\}}, u_{\{2\}}, u_{\{3\}}, u_{\{1,2\}}, u_{\{2,3\}}, u_{\{1,3\}}, u_{\{1,2,3\}}) \in \Gamma_{FSA}(3), \\ & u_{\{1\}} = a, u_{\{2\}} = b, u_{\{3\}} = c, \\ & u_{\{1,2\}} = a + b, u_{\{2,3\}} = b + c, u_{\{1,3\}} = c + a \} \end{aligned}$$

Then $R_{a,b,c}$ is the singleton containing the number $a + b + c$; this follows from the equality conditions for the entropy power inequality, which mandate based on the defining conditions of $R_{a,b,c}$ that X_1, X_2 and X_3 are Gaussian with proportional covariance matrices. On the other hand, $R'_{a,b,c}$ is $\{u_{\{1,2,3\}} \in \mathbb{R}_+ : u_{\{1,2,3\}} \geq a + b + c\}$. This implies that $\overline{\Gamma(3)}$ is a strict subset of $\Gamma_{FSA}(3)$.

²Indeed, one can force $h(X + X')$ to be infinite; note that $h(X + X') < \infty$ if and only if

$$\sum_{n \in \mathbb{N}} a_n \log \frac{1}{a_n} < \infty.$$

This was what motivated [7] to consider this construction, since they wished to exhibit a density f with finite differential entropy whose self-convolution has infinite differential entropy.

F. Proof of Theorem 6

It is easy to see from the definitions that $\overline{\Gamma(n)} \supset \Gamma_\infty(n)$.

We wish to show now that $\overline{\Gamma(n)} \subset \Gamma_\infty(n)$. Let $a \in \overline{\Gamma(n)}$. Then there exists a sequence $a_m \in \Gamma_{d_m}(n)$ such that $a_m \rightarrow a$ in \mathbb{R}^{2^n-1} ; let us write $a_m(s) = e^{2h(\sum_{i \in s} X_{m,i})/d_m}$, where each $X_{m,i}$ is a d_m -dimensional random vector. We may always assume, without loss of generality, that the sequence $d_m \rightarrow \infty$ as $m \rightarrow \infty$. This is because $\Gamma_d(n)$ is trivially embedded in $\Gamma_{ld}(n)$ for any positive integer l ; so we can artificially enforce the condition that $d_m \geq m$ for each m in order to make ∞ the unique limit point of the sequence (d_m) .

For each $k \in \mathbb{N}$, set $D_k = \prod_{m=1}^k d_m$, and consider the random vector Z obtained by stacking 1 copy of X_1 , $D_1 = d_1$ copies of X_2 , $D_2 = d_1 d_2$ copies of X_3 , and so on (to infinity). Let us call Z_k the result of this process if we stopped at the k -th step. Note that Z_k is a random matrix with n columns, and $\mathcal{D}_k := d_1 + D_1 d_2 + \dots + D_{k-1} d_k = \sum_{m=1}^k D_m$ rows of real-valued random variables. We claim that Z is a stochastic process whose point in the infinite-dimensional Stam region is a , which would prove our desired result, and devote the rest of this section to proving this claim.

Denote by A_k the point in the Stam region determined by Z_k . Then

$$A_k(s) = \prod_{m=1}^k N\left(\sum_{i \in s} Z_{m,i}\right)^{D_m/\mathcal{D}_k} = \prod_{m=1}^k a_m(s)^{D_m/\mathcal{D}_k},$$

so that

$$\log A_k(s) = \sum_{m=1}^k \frac{D_m}{\mathcal{D}_k} \log a_m(s).$$

By a modification of the basic theorem about Cesàro means (see, e.g., [22]) combined with the continuity of the exponential and logarithmic functions, it follows from $a_m \rightarrow a$ as $m \rightarrow \infty$ that $A_k \rightarrow a$ as $k \rightarrow \infty$. In particular, since the entropy power rate of $\sum_{i \in s} Z_i$ is just the limit of entropy powers of $\sum_{i \in s} Z_{k,i}$ by definition, we have that the entropy power rate of $\sum_{i \in s} Z_i$ is $a(s)$, and consequently that the point a is in $\Gamma_\infty(n)$.

IV. DISCUSSION OF A QUESTION

Let us start by proving a general proposition about determinants of sums of positive-definite matrices.

Proposition 3. *Let M_1, M_2, \dots, M_n be $d \times d$ positive-definite matrices. Then, the function $v_c : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ defined by*

$$v_c(x) = \det\left(\sum_{i=1}^n x_i M_i\right)$$

is supermodular, i.e. for any $x, y \in \mathbb{R}_+^n$ it holds that

$$v_c(x) + v_c(y) \leq v_c(x \vee y) + v_c(x \wedge y) \quad (11)$$

where $x \vee y$ denotes the componentwise maximum of x and y and $x \wedge y$ denotes the componentwise minimum of x and y .

Proof. Inequality (11) trivially holds when $x = 0$ or $y = 0$. To prove that (11) also holds otherwise, it suffices to show by Proposition 2 (since v_c is smooth) that $\frac{\partial^2 v_c(x)}{\partial x_j \partial x_i} \geq 0$ for any $x \in \mathbb{R}_+^n$. We note that, setting $M = \sum_{i=1}^n x_i M_i$,

$$\frac{\partial v_c(x)}{\partial x_i} = \det(M) \operatorname{tr}(M^{-1} M_i)$$

and

$$\frac{\partial^2 v_c(x)}{\partial x_j \partial x_i} = \det(M) [\operatorname{tr}(M^{-1} M_i) \operatorname{tr}(M^{-1} M_j) - \operatorname{tr}(M^{-1} M_j M^{-1} M_i)]. \quad (12)$$

However, for any two positive definite matrices A and B , $\det(A) \geq 0$ and $\operatorname{trace}(A)\operatorname{trace}(B) \geq \operatorname{trace}(AB)$ (see e.g., Zhang [16], page 166). Hence, both terms on the right-hand side of (12) are always non-negative. \square

By Lemma 1, Proposition 3 immediately yields:

Theorem 7. *For any positive-definite matrices A, B and C , we have*

$$\det(A + B + C) + \det(A) \geq \det(A + B) + \det(A + C). \quad (13)$$

In other words, the set function $s \mapsto e^{2h(\sum_{i \in s} X_i)}$ is supermodular if all the random vectors X_i are Gaussian (since in this case, the functional e^{2h} is—up to irrelevant constants—the determinant of the covariance matrix).

It is natural to wonder how far this phenomenon extends. For example, one might ask if it holds for arbitrary distributions of the X_i , which— if true— would imply that $\Gamma_1(n) \subset \Gamma_{SM}(n)$, thus refining the inclusion $\Gamma_1(n) \subset \Gamma_{FSA}(n)$ that follows from Theorem 1. However, this turns out to be false. Indeed, we will see that even a much more restrictive statement is false, but to discuss it, we need to take a detour through some convex geometry.

It was recently shown in [15] (in two different ways, one of which combines the determinant inequality of Theorem 7 with some tools from optimal transportation theory) that the supermodularity property of Theorem 7 can be extended from determinants to volumes of convex sets.

Theorem 8. [15] *If B_i are convex sets in \mathbb{R}^d ,*

$$\text{Vol}_d(B_1 + B_2 + B_3) + \text{Vol}_d(B_1) \geq \text{Vol}_d(B_1 + B_2) + \text{Vol}_d(B_1 + B_3). \quad (14)$$

Theorem 8 is a generalization of Theorem 7 since volumes of ellipsoids and parallelotopes (which are all convex) are given by determinants, and since Minkowski summation of these objects correspond to addition of the corresponding positive-definite matrices. It is observed in [15] that Theorem 8 does not extend to arbitrary compact sets; however, the question is posed there (and verified in dimension 1) whether the inequality (14) holds when B_1 is compact and convex, while B_2, B_3 are arbitrary compact sets.

As is now well known, the entropy power inequality $\mathcal{N}(X + Y) \geq \mathcal{N}(X) + \mathcal{N}(Y)$ resembles in many ways the Brunn-Minkowski inequality, which is a very important inequality in mathematics, and a cornerstone of convex geometry in particular. This was first noticed by Costa and Cover [12] (see also [13], [48], [51] for other aspects of this connection). The analogies between inequalities in information theory and those in convex geometry have been explored quite a bit in recent years (see, e.g., the survey [36] and references therein), based loosely on the understanding that the functional $A \mapsto \text{Vol}_d(A)^{1/d}$ in the geometry of compact subsets of \mathbb{R}^d , and the functional $\mathcal{L}(X) \mapsto \mathcal{N}(X)$ in probability are analogous to each other in many ways. In the dictionary that relates notions in convex geometry to those in probability, the natural analog of a convex set is a log-concave distribution. Recall that a log-concave distribution is one that has a density of the form e^{-V} , where $V : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ is a convex function.

Motivated by the afore-mentioned question of [15], it is tempting to consider the following question which would be its natural probabilistic analog:

Question 1. *If X, Y, Z are independent \mathbb{R}^d -valued random vectors with Z having a log-concave distribution, is it true that*

$$e^{2h(X+Y+Z)} + e^{2h(Z)} \geq e^{2h(X+Z)} + e^{2h(Y+Z)}?$$

If the answer to Question 1 is positive, it would imply that for independent \mathbb{R}^d -valued random vectors X_1, \dots, X_n with log-concave distributions, the set function

$$v(s) = e^{2h(\sum_{i \in s} X_i)}, \quad s \subset [n]$$

is supermodular. Unfortunately, however, we will now show that the answer to Question 1 is negative.

Proposition 4. *There exist independent \mathbb{R} -valued random variables X, Y, Z , each with log-concave distributions, such that*

$$e^{2h(X+Y+Z)} + e^{2h(Z)} < e^{2h(X+Z)} + e^{2h(Y+Z)}?$$

In particular, the answer to Question 1 is negative, already in dimension 1.

Proof. Suppose the answer to Question 1 were yes for $d = 1$. Then, given any $\epsilon > 0$, taking X_ϵ to be Gaussian with variance ϵ would yield

$$\mathcal{N}(X_\epsilon + Y + Z) + \mathcal{N}(Z) \geq \mathcal{N}(X_\epsilon + Z) + \mathcal{N}(Y + Z).$$

On rearrangement, we have

$$\mathcal{N}(X_\epsilon + Y + Z) - \mathcal{N}(Y + Z) \geq \mathcal{N}(X_\epsilon + Z) - \mathcal{N}(Z).$$

Dividing by ϵ and taking the limit as $\epsilon \rightarrow 0$ gives

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \mathcal{N}(X_\epsilon + Y + Z) \geq \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \mathcal{N}(X_\epsilon + Z). \quad (15)$$

The classical de Bruijn identity [47], which is easily proved by observing that the density of $X_\epsilon + U$ satisfies the heat equation when X_ϵ is a Gaussian of variance ϵ independent of U , asserts that

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} h(X_\epsilon + U) = I(U),$$

where I is the Fisher information, i.e., if U has density u ,

$$I(U) = \int_{\mathbb{R}} \frac{u'(x)^2}{u(x)} dx.$$

As observed by Stam [47], the chain rule implies that

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \mathcal{N}(X_\epsilon + U) = \mathcal{N}(U)I(U). \quad (16)$$

Putting together the inequality (15) and the identity (16), we conclude that³

$$\mathcal{N}(X + Y)I(X + Y) \geq \mathcal{N}(X)I(X).$$

Setting $T_n = \sum_{i=1}^n X_i$ and $S_n = \frac{1}{\sqrt{n}}T_n$ for i.i.d. random variables X_i with the same distribution as X , we must have

$$\mathcal{N}(S_n)I(S_n) = \mathcal{N}(T_n)I(T_n) \geq \mathcal{N}(T_{n-1})I(T_{n-1}) = \mathcal{N}(S_{n-1})I(S_{n-1}),$$

which upon taking the limit as $n \rightarrow \infty$ and using the entropic central limit theorem of Barron [3] and the corresponding statement for Fisher information [24, Theorem 1.6], yields the inequality $2\pi e \geq \mathcal{N}(X)I(X)$. Since this contradicts the entropic isoperimetric inequality, the assumption that we started with—namely, that the answer to Question 1 is yes for $d = 1$ —cannot hold. \square

Since the answer to Question 1 is negative in dimension 1, whereas the corresponding question for volumes of sets has a positive answer in dimension 1 as observed in [15], Proposition 4 adds to the examples where the analogy between the probabilistic and convex geometric statements breaks down.

Proposition 4 may also be seen as a strengthening of Theorem 2. While the proof of Theorem 2 in Section III-B shows that $\Gamma_d(n) \cap \Gamma_{SM}(n)^c \neq \emptyset$ for $d > 1$ and $n \geq 3$, Proposition 4 shows that for $n \geq 3$, even $\Gamma_1(n)$ contains points outside of the supermodularity cone $\Gamma_{SM}(n)$. Indeed, by examining the proof, it shows something much stronger: even if we restrict ourselves to the class of one-dimensional random variables X, Y, Z with Z being Gaussian and X, Y being i.i.d. and drawn from a log-concave distribution, supermodularity of entropy power fails.

V. REMARKS

- 1) The classical entropy power inequality of Shannon [43] and Stam [47] states that if X_i are independent \mathbb{R}^d -valued random vectors,

$$\mathcal{N}(X_1 + \dots + X_n) \geq \sum_{j=1}^n \mathcal{N}(X_j). \quad (17)$$

Motivated by the long-standing monotonicity conjecture for the entropic central limit theorem, Artstein, Ball, Barthe and Naor [1] proved a new entropy power inequality:

$$\mathcal{N}(X_1 + \dots + X_n) \geq \frac{1}{n-1} \sum_{i=1}^n \mathcal{N}\left(\sum_{j \neq i} X_j\right). \quad (18)$$

Simplified proofs of inequality (18) were independently given by [29], [50], [46]. Subsequently, [30] proved the inequality (7) based on the maximum degree of the hypergraph \mathcal{G} ; this contains the inequalities (18) and (17) since for $\mathcal{G} = \mathcal{G}_m$, namely the collection of all subsets of indices of size m , we have $r = \binom{n-1}{m-1}$. Theorem 1 says that for any fractional partition β using a collection \mathcal{G} of subsets of $[n]$,

$$\mathcal{N}(X_1 + \dots + X_n) \geq \sum_{s \in \mathcal{G}} \beta_s \mathcal{N}\left(\sum_{j \in s} X_j\right). \quad (19)$$

Since coefficients that are identical to the reciprocal of the maximum degree of a hypergraph constitute a fractional partition, Theorem 1 subsumes and extends all the inequalities discussed above.

- 2) What are the basic properties of the Stam region? Apart from the theorems stated in the introduction, we do not know any other structural properties of $\Gamma_d(n)$ or $\Gamma(n)$. There are many natural questions, such as whether $\Gamma(n)$ or its closure is convex, that remain to be answered.

³Since $\mathcal{N}(X + Y) \geq \mathcal{N}(X)$ and $I(X + Y) \leq I(X)$, if such an inequality were true, it would capture how \mathcal{N} and I compete in their behavior on convolution in the log-concave setting. Observe that $\mathcal{N}(X)I(X)$ is an affine-invariant functional in dimension 1, and that a limiting argument based on the entropy power inequality (see, e.g., [47]) yields the entropic isoperimetric inequality

$$\mathcal{N}(X)I(X) \geq 2\pi e,$$

with equality if and only if X is Gaussian.

- 3) There are additional constraints on the Stam region for $n \geq 3$ that have not been discussed so far, but these are nonlinear. Specifically, it was observed in [28] (see [34], [35], [5], [6], [9], [26], [32], [33] for various applications and generalizations) that the differential entropy of sums of independent random vectors is submodular, i.e.,

$$h(X_1 + X_2 + X_3) + h(X_1) \leq h(X_1 + X_2) + h(X_1 + X_3).$$

Written in terms of entropy powers, this implies

$$N(X_1 + X_2 + X_3)N(X_1) \leq N(X_1 + X_2)N(X_1 + X_3), \quad (20)$$

which is a nonlinear constraint on the Stam region for any $n \geq 3$.

- 4) The study of the Stam region is analogous in some sense to the study of the entropic region defined using the joint entropy of subsets of random variables, on which there has been much progress in recent decades. Let X_s denote $(X_i : i \in s)$ and H denote the discrete entropy, where $(X_i : i \in [n])$ is a collection of *dependent* random variables taking values in some finite set. Fujishige [17] observed that $g(s) = H(X_s)$ is a submodular set function. Consequent entropy inequalities obtained by Han [21] and Shearer [10] became influential in information theory and in combinatorics respectively. These inequalities were unified and generalized in [39], [37], where it was shown that for any submodular set function $g : 2^{[n]} \rightarrow \mathbb{R}_+$ (and in particular for the joint entropy), one has

$$g([n]) \leq \sum_{s \in \mathcal{G}} \beta_s g(s), \quad (21)$$

as well as corresponding lower bounds that we do not state here, for any fractional partition β using any collection of sets \mathcal{G} . Motivated by the problem of characterizing the entropic region (or equivalently, the class of all entropy inequalities for the joint distributions of a collection of dependent random variables), Zhang and Yeung [56] observed in 1998 that there exist so-called non-Shannon inequalities that do not follow from submodularity of joint entropy; these have seen much active investigation recently (see, e.g., Matúš [38], who showed the remarkable fact that the entropic region is not polyhedral if $n \geq 4$). While Theorem 1 is analogous in some sense to the inequality (21), Theorem 2 shows that the analogue of Fujishige's submodularity is not true. In particular, the question of whether there exist entropy inequalities for sums that are formally analogous to non-Shannon inequalities is ill-posed, since we do not have supermodularity of entropy power to start with.

- 5) Theorem 1 is a more informative statement than its predecessors such as (7), as pointed out in [27]. Recall that entropy power inequalities have been key to the determination of some capacity and rate regions, and that rate regions for several multi-user problems (such as the m -user Slepian-Wolf problem) involve subset sum constraints. Vaguely motivated by this, one may consider the “region” of all $(R_1, \dots, R_n) \in \mathbb{R}_+^n$ satisfying $\sum_{j \in s} R_j \geq \mathcal{N}(T^s)$ for each $s \subset [n]$. Then Theorem 1 is equivalent to the existence of a point in this region such that the total sum $\sum_{j \in [n]} R_j = \mathcal{N}(T^{[n]})$. Although we are not yet aware of a specific multiuser capacity problem with precisely this rate region, this fact appears intriguing.
- 6) As discussed in Section IV, there are extensive analogies between inequalities in convex geometry and entropy inequalities. The volume analog of the fractional entropy power inequalities proved in this paper, namely,

$$\text{Vol}_d^{1/d} \left(\sum_{i=1}^n A_i \right) \geq \frac{1}{n-1} \sum_{i=1}^n \text{Vol}_d^{1/d} \left(\sum_{j \in [n] \setminus \{i\}} A_j \right),$$

was observed to hold for Minkowski sums of compact convex sets in [8], and conjectured to hold more generally for Minkowski sums of compact sets. However, recently it was shown in [14] that this conjecture for general compact sets fails in dimension 12 or greater. (On the other hand, it was shown in [15] that the fractional superadditivity inequality

$$\text{Vol}_d \left(\sum_{i=1}^n A_i \right) \geq \frac{1}{n-1} \sum_{i=1}^n \text{Vol}_d \left(\sum_{j \in [n] \setminus \{i\}} A_j \right),$$

without the exponents applied to the volumes, holds for general compact sets A_i .)

- 7) While this paper focused on characterization of possible inequalities for the entropies of sums of independent random vectors in \mathbb{R}^d , the same question also makes sense (and is interesting both as a basic mathematical question and in view of applications to communication theory) for random variables taking values in any group. A priori, it is natural to first try groups that have particularly simple structure—such as finite cyclic groups or the integers. However, even for these seemingly staid examples, there is little that can currently be said. Indeed, we do not even know a fully satisfactory analogue of the entropy power inequality on the integers (some partial results are available in [23], [45], [20], [52], [53]). As in the Euclidean setting, such inequalities for integer-valued random variables also have connections to probabilistic limit theorems (see, e.g., [2], [54], [25]).

ACKNOWLEDGMENT

The first author thanks Young-Han Kim for suggesting that this note, which was largely written in 2009 but set aside, may be of broader interest and for the encouragement to complete and submit it for publication. Parts of this work were completed while the first author was at the Institute of Mathematics and its Applications (IMA) in Minneapolis during spring 2015, and he is grateful to the IMA for its hospitality and providing a conducive research environment.

REFERENCES

- [1] S. Artstein, K. M. Ball, F. Barthe, and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.*, 17(4):975–982 (electronic), 2004.
- [2] A. D. Barbour, O. Johnson, I. Kontoyiannis, and M. Madiman. Compound Poisson approximation via information functionals. *Electron. J. Probab.*, 15(42):1344–1368, 2010.
- [3] A.R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14:336–342, 1986.
- [4] P. Bergmans. A simple converse for broadcast channels with additive white gaussian noise. *IEEE Trans. Inform. Theory*, 20(2):279–280, 1974.
- [5] S. Bobkov and M. Madiman. Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.*, 349:201–204, Février 2011.
- [6] S. Bobkov and M. Madiman. Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures. *J. Funct. Anal.*, 262:3309–3339, 2012.
- [7] S. G. Bobkov and G. P. Chistyakov. Entropy power inequality for the Rényi entropy. *IEEE Trans. Inform. Theory*, 61(2):708–714, February 2015.
- [8] S. G. Bobkov, M. Madiman, and L. Wang. Fractional generalizations of Young and Brunn-Minkowski inequalities. In C. Houdré, M. Ledoux, E. Milman, and M. Milman, editors, *Concentration, Functional Inequalities and Isoperimetry*, volume 545 of *Contemp. Math.*, pages 35–53. Amer. Math. Soc., 2011.
- [9] S. G. Bobkov and M. M. Madiman. On the problem of reversibility of the entropy power inequality. In *Limit theorems in probability, statistics and number theory*, volume 42 of *Springer Proc. Math. Stat.*, pages 61–74. Springer, Heidelberg, 2013. Available online at [arXiv:1111.6807](https://arxiv.org/abs/1111.6807).
- [10] F.R.K. Chung, R.L. Graham, P. Frankl, and J.B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Combinatorial Theory, Ser. A*, 43:23–37, 1986.
- [11] M.H.M. Costa. On the Gaussian interference channel. *IEEE Trans. Inform. Theory*, 31(5):607–615, 1985.
- [12] M.H.M. Costa and T.M. Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. *IEEE Trans. Inform. Theory*, 30(6):837–839, 1984.
- [13] A. Dembo, T.M. Cover, and J.A. Thomas. Information-theoretic inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, 1991.
- [14] M. Fradelizi, M. Madiman, A. Marsiglietti, and A. Zvavitch. Do Minkowski averages get progressively more convex? *C. R. Acad. Sci. Paris Sér. I Math.*, 354(2):185–189, February 2016.
- [15] M. Fradelizi, M. Madiman, A. Marsiglietti, and A. Zvavitch. On the monotonicity of Minkowski sums towards convexity. *Preprint*, 2017.
- [16] E. Friedgut and J. Kahn. On the number of copies of one hypergraph in another. *Israel Journal of Mathematics*, 105:251–256, 1998.
- [17] S. Fujishige. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39:55–72, 1978.
- [18] R. Gill, M. van der Laan, and J. Robins. Coarsening at random: Characterisations, conjectures and counter-examples. In D. Lin, editor, *Proceedings First Seattle Conference on Biostatistics*, pages 255–294. Springer, New York, 1997.
- [19] R. D. Gill and P. D. Grünwald. An algorithmic and a geometric characterization of coarsening at random. *Ann. Statist.*, 36(5):2409–2422, 2008.
- [20] S. Haghshatoor, E. Abbe, and E. Telatar. A new entropy power inequality for integer-valued random variables. *IEEE Trans. Inform. Th.*, 60(7):3787–3796, July 2014.
- [21] Te Sun Han. Nonnegative entropy measures of multivariate symmetric correlations. *Information and Control*, 36(2):133–156, 1978.
- [22] G. H. Hardy. *Divergent series*. Éditions Jacques Gabay, Sceaux, 1992. With a preface by J. E. Littlewood and a note by L. S. Bosanquet, Reprint of the revised (1963) edition.
- [23] P. Harremoës and C. Vignat. An entropy power inequality for the binomial family. *J. Inequal. Pure Appl. Math.*, 4(5):Article 93, 6 pp. (electronic), 2003.
- [24] O. Johnson and A.R. Barron. Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, 129(3):391–409, 2004.
- [25] O. Johnson, I. Kontoyiannis, and M. Madiman. Log-concavity, ultra-log-concavity, and a maximum entropy property of discrete compound Poisson measures. *Discrete Appl. Math.*, 161:1232–1250, 2013. DOI: 10.1016/j.dam.2011.08.025.
- [26] I. Kontoyiannis and M. Madiman. Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inform. Theory*, 60(8):4503–4514, August 2014.
- [27] M. Madiman. Cores of cooperative games in information theory. *EURASIP J. on Wireless Comm. and Networking*, (318704), 2008.
- [28] M. Madiman. On the entropy of sums. In *Proc. IEEE Inform. Theory Workshop*, pages 303–307. Porto, Portugal, 2008.
- [29] M. Madiman and A.R. Barron. The monotonicity of information in the central limit theorem and entropy power inequalities. In *Proc. IEEE Intl. Symp. Inform. Theory*, pages 1021–1025. Seattle, July 2006.
- [30] M. Madiman and A.R. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory*, 53(7):2317–2329, July 2007.
- [31] M. Madiman and F. Ghassemi. The entropy power of sums is fractionally superadditive. In *Proc. IEEE Intl. Symp. Inform. Theory*, pages 295–298. Seoul, Korea, 2009.
- [32] M. Madiman and I. Kontoyiannis. The entropies of the sum and the difference of two IID random variables are not too different. In *Proc. IEEE Intl. Symp. Inform. Theory*, Austin, Texas, June 2010.
- [33] M. Madiman and I. Kontoyiannis. Entropy bounds on abelian groups and the Ruzsa divergence. *Preprint*, [arXiv:1508.04089](https://arxiv.org/abs/1508.04089), 2015.
- [34] M. Madiman, A. Marcus, and P. Tetali. Information-theoretic inequalities in additive combinatorics. In *Proc. IEEE Inform. Theory Workshop*, Cairo, Egypt, January 2010.
- [35] M. Madiman, A. Marcus, and P. Tetali. Entropy and set cardinality inequalities for partition-determined functions. *Random Struct. Alg.*, 40:399–424, 2012.
- [36] M. Madiman, J. Melbourne, and P. Xu. Forward and reverse entropy power inequalities in convex geometry. In E. A. Carlen, M. Madiman, and E. Werner, editors, *Convexity and Concentration*, IMA Volumes in Mathematics and its Applications. Springer, to appear. Available online at [arXiv:1604.04225](https://arxiv.org/abs/1604.04225).
- [37] M. Madiman and P. Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Inform. Theory*, 56(6):2699–2713, June 2010.
- [38] F. Matúš. Infinitely many information inequalities. In *Proc. IEEE Intl. Symp. Inform. Theory*, pages 41–44, Nice, France, 2007.
- [39] J. Moulin Ollagnier and D. Pinchon. Filtre moyennant et valeurs moyennes des capacités invariantes. *Bull. Soc. Math. France*, 110(3):259–277, 1982.
- [40] Y. Oohama. The rate-distortion function for the quadratic gaussian cto problem. *IEEE Trans. Inform. Theory*, 44(3):1057–1070, 1998.
- [41] L. Ozarow. On a source coding problem with two channels and three receivers. *Bell System Tech. J.*, 59(10):1909–1921, 1980.
- [42] J. L. Ramírez Alfonsín. *The Diophantine Frobenius problem*, volume 30 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2005.
- [43] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

- [44] L. S. Shapley. On balanced sets and cores. *Naval Research Logistics Quarterly*, 14:453–560, 1967.
- [45] N. Sharma, S. Das, and S. Muthukrishnan. Entropy power inequality for a family of discrete random variables. In *Proc. IEEE Intl. Symp. Inform. Theory.*, pages 1945–1949. St. Petersburg, 2011.
- [46] D. Shlyakhtenko. A free analogue of Shannon’s problem on monotonicity of entropy. *Adv. Math.*, 208(2):824–833, 2007.
- [47] A.J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.
- [48] S. J. Szarek and D. Voiculescu. Shannon’s entropy power inequality via restricted Minkowski sums. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 257–262. Springer, Berlin, 2000.
- [49] D. M. Topkis. *Supermodularity and complementarity*. Frontiers of Economic Research. Princeton University Press, Princeton, NJ, 1998.
- [50] A. M. Tulino and S. Verdú. Monotonic decrease of the non-gaussianity of the sum of independent random variables: A simple proof. *IEEE Trans. Inform. Theory*, 52(9):4295–7, September 2006.
- [51] L. Wang and M. Madiman. Beyond the entropy power inequality, via rearrangements. *IEEE Trans. Inform. Theory*, 60(9):5116–5137, September 2014.
- [52] L. Wang, J. O. Woo, and M. Madiman. A lower bound on the Rényi entropy of convolutions in the integers. In *Proc. IEEE Intl. Symp. Inform. Theory*, pages 2829–2833. Honolulu, Hawaii, July 2014.
- [53] J. O. Woo and M. Madiman. A discrete entropy power inequality for uniform distributions. In *Proc. IEEE Intl. Symp. Inform. Theory*, Hong Kong, China, June 2015.
- [54] Y. Yu. Monotonic convergence in an information-theoretic law of small numbers. *IEEE Trans. Inform. Theory*, 55(12):5412–5422, 2009.
- [55] F. Zhang. *Matrix theory: Basic results and techniques*. Springer-Verlag, New York, 1999.
- [56] J. Zhang and R.W. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory*, 44:1440–1452, 1998.